DOCUMENT RESUME

ED 391 813 TM 024 152

AUTHOR	Lord, Frederic M.; Pashley, Peter J.
TITLE	Confidence Bands for the Three-Parameter Logistic Item Response Curve.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-88-67
PUB DATE	Dec 88
NOTE	23p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer Software Development; Goodness of Fit; *Item Response Theory; *Maximum Likelihood Statistics; *Sampling
IDENTIFIERS	Calibration; *Confidence Intervals (Statistics); *Three Parameter Model

ABSTRACT

A large sample method for obtaining asymptotic simultaneous confidence bands for a three-parameter logistic response curve is described. Simultaneous confidence bands indicate the sampling variation of item response curves relative to a fitted function. A procedure is given which requires as input maximum likelihood parameter estimates and an asymptotic error covariance matrix. Illustrative plots of item response curves with corresponding confidence bands are presented. These procedures are simple enough computationally to be used economically in regular item response studies. Incorporating the calibration and graphical display of simultaneous confidence bands into calibration software would greatly benefit test designers. (Contains four figures and seven references.) (Author/SLD)

RESEARCH REPORT

CONFIDENCE BANDS FOR THE THREE-PARAMETER LOGISTIC ITEM RESPONSE CURVE

Frederic M. Lord
Peter J. Pashley

Educational Testing Service
Princeton, New Jersey
December 1988

Confidence Bands for the Three-Parameter Logistic

Item Response Curve

Frederic M. Lord and Peter J. Pashley

Educational Testing Service

November 1988

Confidence Bands for the Three-Parameter Logistic Item Response Curve

Abstract

A large sample method for obtaining asymptotic simultaneous confidence bands for a three-parameter logistic response curve is described. A procedure is given which requires as input maximum likelihood parameter estimates and an asymptotic error covariance matrix. Illustrative plots of item response curves with corresponding confidence bands are presented.

Confidence Bands for the Three-Parameter Logistic Item Response Curve

Introduction

The three-parameter logistic (Birnbaum) item response function is commonly used to model latent abilities in psychometric studies. In most situations maximum likelihood parameter estimates are obtainable through computer packages such as LOGIST (Wingersky, 1985; Wingersky, Barton, & Lord, 1982). Unfortunately, simultaneous confidence bands for the resulting item response curves have not been available to psychometricians.

A graphical approach using N-line plots for the three-parameter model has been suggested by Thissen and Wainer (1983a, 1983b). This technique is convenient for determining the density of curves in the confidence envelope corresponding to any monotonic function. However, the number of curves needed to obtain 95% confidence bands for the three-parameter logistic model is 235, which is far too many to plot in most practical situations.

Hauck (1983) developed a procedure for obtaining confidence bands for any model whose logit is linear in the parameters. This method is similar to those of linear regression models and may be applied to the two-parameter logistic response function. Unfortunately, the addition of a guessing parameter negates the applicability of Hauck's approach to the three-parameter model.

A procedure specifically for determining large sample confidence bands for the two-parameter logistic model was first developed by

Brand, Pinnock, and Jackson (1973). Their method involved solving algebraically for extrema on the confidence ellipse for the logit model parameters. Obtained minima and maxima, for various latent ability values, may then be used to form simultaneous confidence bands for the response curve.

A similar approach is taken here for the three-parameter logistic model. While explicit formulas for the confidence regions are not available in this case, a convenient numerical procedure is presented. Examples of confidence bands for SAT test items are given.

Background

The form of the logistic function used throughout this paper is

$$P(\theta) = c + \frac{1 - c}{1 + e^{-L}}, \qquad 0 \leq c \leq 1, \tag{1}$$

where $L = A\theta + B$, θ is the usual ability parameter, and A, and B and c are item parameters. See the Appendix for the relationship between this equation and the more common formulation given by Lord (1980).

For a given item, the asymptotic distribution of the maximum likelihood estimator of the item parameters $\hat{\mu} = (\hat{A}, \hat{B}, \hat{c})'$, assuming the θ_a ($a = 1, 2, \ldots, N$) to be known, is the normal trivariate distribution

$$(2\pi)^{-3/2} |\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2} (\mu - \hat{\mu})' \mathbf{V}^{-1} (\mu - \hat{\mu}) \right] \tag{2}$$

where $\mu = (A,B,c)'$ and $\mathbf{V}^{-1}$ is the Fisher information matrix. Since the quadratic form in the exponent of a normal distribution is distributed as chi-square, in repeated sampling, asymptotically

$$\mathrm{Prob}[(\mu - \hat{\mu})'\mathbf{V}^{-1}(\mu - \hat{\mu}) \leq \chi^2_{1-\alpha} \mid \mu, \mathbf{V}] = 1-\alpha \tag{3}$$

where $\chi^2_{1-\alpha}$ is the $100(1-\alpha)$ -th percentile of the chi-square distribution with three degrees of freedom.

The matrix $\mathbf{V}$ may be replaced by $\hat{\mathbf{V}}$ without changing the asymptotic validity of (3). An asymptotic $(1-\alpha)$ confidence region for μ may be computed from observed sample values $\hat{\mu}$ and $\hat{\mathbf{V}}$:

$$(\mu - \hat{\mu})'\hat{\mathbf{V}}^{-1}(\mu - \hat{\mu}) \leq \chi^2_{1-\alpha} . \tag{4}$$

For given θ, let $P_L(\theta)$ and $P_U(\theta)$ denote the minimum and maximum values of $P(\theta)$ found within the ellipsoid region (4). Now consider $P_L(\theta)$ and $P_U(\theta)$ as functions of θ, defining a confidence band within which the true $P(\theta)$ is asserted to lie.

Consider drawing repeated samples. From each sample, construct an ellipsoid region (4). In $1-\alpha$ of the samples, the ellipsoid will contain the true parameter value μ. Corresponding to each μ is the true item response function $P(\theta)$. When μ lies within the ellipsoid (4)

$$P_L(\theta) \leq P(\theta) \leq P_U(\theta) . \tag{5}$$

Since μ lies within the ellipsoid in $1-\alpha$ of all the samples, the inequality (5) will hold for at least 1-α of all samples. Thus the assertion that (5) holds will be correct at least 1-α of the time in repeated sampling. The words 'at least' are necessary here because (5) may also hold for some of the samples in which μ lies outside the ellipsoid. From this result follows the 'simultaneous' property that in 1 - α of the samples the entire curve will fall between P_L and P_U.

It remains to find for each θ the maximum and minimum values of $P(\theta)$ within the ellipsoid (4). Previous attempts to do this [by F. M. Lord (personal communication), and by Thissen and Wainer (1983a, 1983b)] found maxima and minima of $P(\theta)$ on the boundary of the ellipsoid region. M. L. Stocking (personal communication), however, demonstrated that these maxima and minima were sometimes local rather than global. It remains to find the largest local maximum and the smallest local minimum.

Reformulating the Optimization Problem

In order to simplify the constraint space given in (4), consider the following. Define

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ \theta & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} , \qquad \mathbf{M}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -\theta & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} . \tag{6}$$

Let

$$\nu = \mathbf{M}\mu = \mathbf{M} \begin{bmatrix} A \\ B \\ c \end{bmatrix} = \begin{bmatrix} A \\ A\theta + B \\ c \end{bmatrix} = \begin{bmatrix} A \\ L \\ c \end{bmatrix} . \tag{7}$$

Thus

$$\mu'\hat{V}^{-1}\mu = \mu'M'M'^{-1}\hat{V}^{-1}M^{-1}M\mu = \nu'\hat{S}^{-1}\nu \tag{8}$$

where $\hat{S}^{-1} = M'^{-1}\hat{V}^{-1}M^{-1}$. Similarly (4) may be written

$$(\nu - \hat{\nu})' \hat{S}^{-1}(\nu - \hat{\nu}) \leq \chi^2_{1-\alpha} \quad . \tag{9}$$

This represents a change in coordinate axes, reexpressing the ellipsoid region (4) in A, B, c as an ellipsoid region (9) in A, L, c for a given θ.

For fixed θ, consider how the item response function (1) varies with L and c. The first partial derivatives are

$$\frac{\partial P(\theta)}{\partial L} = \frac{1 - c}{(1 + e^{L})(1 + e^{-L})} \ , \tag{10}$$

and

$$\frac{\partial P(\theta)}{\partial c} = \frac{1}{1 + e^{L}} \ . \tag{11}$$

Since both of these derivatives are always positive, within any given ellipsoid both the minimum and the maximum values of $P(\theta)$ for any fixed θ will be found on the boundary of the ellipsoid. That is, the constraint space (9) may be reduced to the ellipsoid

$$(\nu - \hat{\nu})'\hat{S}^{-1}(\nu - \hat{\nu}) = \chi^2_{1-\alpha} \quad . \tag{12}$$

Numerical Solution

For fixed θ, the problem is to find the values of A, L, and c satisfying (12) that yield the extrema of $P(\theta)$. This nonlinear maximization problem (with quadratic constraints) can be solved using Lagrange multiplier method. The Lagrange function is

$$P(\theta) - \frac{1}{2} \lambda \left(\sum_{i}^{3} \sum_{j}^{3} v_i v_j t_{ij} - \chi^2_{1-\alpha} \right) \tag{13}$$

where $-\frac{1}{2} \lambda$ is the Lagrange multiplier, $v_i = \nu_i - \hat{\nu}_i$, and $\|t_{ij}\| = \hat{S}^{-1}$.

The desired extrema of $P(\theta)$ are found by setting equal to zero the derivatives of the Lagrange function with respect to A, L, and c. Using (10) and (11), these derivatives are

$$A: \quad -\lambda \sum_{j}^{3} v_j t_{1j} = 0 , \tag{14}$$

$$B: \quad \frac{(1 - c)r}{1 + e^{L}} - \lambda \sum_{j}^{3} v_j t_{2j} = 0 , \tag{15}$$

and

$$c: \quad \frac{1}{1 + e^{L}} - \lambda \sum_{j}^{3} v_j t_{3j} = 0 , \tag{16}$$

where $r = 1/(1 + e^{-L})$.

To solve for λ, multiply (14) by v_1, (15) by v_2, (16) by v_3 and add to find:

$$\frac{r(1 - c)v_2}{1 + e^{L}} + \frac{v_3}{1 + e^{L}} = \lambda \sum_{i}^{3} \sum_{j}^{3} v_i v_j t_{ij} .$$

By (12) the double sum is $\chi^2_{1-\alpha}$, so

$$\lambda = \frac{r(1 - c)v_2 + c - \hat{c}}{\chi^2_{1-\alpha}(1 + e^L)} = \frac{\alpha_2 c + \beta_2}{1 + e^L} , \tag{17}$$

where

$$\alpha_2 = (1 - rv_2)/\chi^2_{1-\alpha} , \qquad \beta_2 = (rv_2 - \hat{c})/\chi^2_{1-\alpha} .$$

An expression for c as a function of L and known quantities is now developed. From (14),

$$v_1 = -\frac{1}{t_{11}}(v_2 t_{12} + t_{13}c - t_{13}\hat{c}) = \alpha_1 c + \beta_1 \tag{18}$$

where

$$\alpha_1 = -t_{13}/t_{11} , \qquad \beta_1 = (t_{13}\hat{c} - t_{12}v_2)/t_{11} .$$

Substitute (17) and (18) into (16):

$$\begin{aligned} 0 &= 1 - (\alpha_2 c + \beta_2)[t_{31}(\alpha_1 c + \beta_1) + t_{32}v_2 + t_{33}c - t_{33}\hat{c}] \\ &= 1 - (\alpha_2 c + \beta_2)(\alpha_3 c + \beta_3) , \end{aligned}$$

where

$$\alpha_3 = t_{31}\alpha_1 + t_{33} , \qquad \beta_3 = t_{31}\beta_1 + t_{32}v_2 - t_{33}\hat{c} .$$

If α_2 and α_3 are non-zero, solve the quadratic for c:

$$c = \frac{1}{2\alpha_2\alpha_3} \{-\alpha_2\beta_3 - \alpha_3\beta_2 \pm [(\alpha_2\beta_3 - \alpha_3\beta_2)^2 + 4\alpha_2\alpha_3]^{1/2}\} . \tag{19}$$

This expresses c as a function of L and the known quantities $\hat{c}$, t_{ij}, $\chi^2_{1-\alpha}$, and $\hat{L} = \hat{A}\theta + \hat{B}$.

Substitute (17) into (15) to find

$$(1 - c)r - (\alpha_2 c + \beta_2) \sum_j^3 v_j t_{2j} = 0 . \tag{20}$$

By substituting (19) for c in (18), v_1 can be expressed in terms of L and known quantities. Substituting this result, along with (19), into (20) yields an equation in the single unknown L, which will be denoted by $G(L)$. This equation must be solved numerically (for each fixed θ) to find all roots. For each root, compute from (1) and (19) the corresponding $P(\theta)$. The largest and smallest values of $P(\theta)$ so obtained (for fixed θ) are $P_U(\theta)$ and $P_L(\theta)$.

In order to find all the roots of $G(L)$, it is helpful to know the maximum and minimum values of L satisfying (12). Setting $G(L) = 0$, the function can be evaluated (for fixed θ) at a large number of equally spaced values of L covering the range between the maximum and minimum values of L. Any interval in which $G(L)$ changes sign can then be searched, using *regula falsi*, to find the root in that interval.

To find the extrema of L satisfying (12), let

$$F(A,L,c) = \sum_i^3 \sum_j^3 v_i v_j t_{ij} - \chi^2_{1-\alpha} = 0 . \tag{21}$$

Then if (21) is viewed as defining L as a function of A and c,

$$\frac{\partial L}{\partial A} = -\frac{\Sigma v_j t_{1j}}{\Sigma v_j t_{2j}} , \qquad \frac{\partial L}{\partial c} = -\frac{\Sigma v_j t_{3j}}{\Sigma v_j t_{2j}} .$$

Setting these derivatives equal to zero gives

$$v_1 t_{11} + v_2 t_{12} + v_3 t_{13} = 0 ,$$

(22)

and $$v_1 t_{31} + v_2 t_{32} + v_3 t_{33} = 0 .$$

Eliminate v_1 to find

$$v_3 = -\frac{(t_{31} t_{12} - t_{11} t_{32}) v_2}{t_{31}^2 - t_{11} t_{33}} = \alpha_4 v_2 , \tag{23}$$

say. Similarly

$$v_1 = -\frac{(t_{31} t_{23} - t_{33} t_{12}) v_2}{t_{31}^2 - t_{11} t_{33}} = \alpha_5 v_2 . \tag{24}$$

Substitute (22) into (21) to find

$$v_2 \Sigma v_j t_{2j} = \chi^2_{1-\alpha} . \tag{25}$$

Substitute (23) and (24) into (25) to find

$$(\alpha_5 t_{21} + t_{22} + \alpha_4 t_{23}) v_2^2 = \chi^2_{1-\alpha} .$$

Thus the extrema of L satisfying (13) are given by

$$L = \hat{L} \pm [\chi^2_{1-\alpha}/(\alpha_5 t_{21} + t_{22} + \alpha_4 t_{23})]^{1/2} . \tag{26}$$

The required values of $\|V^{rs}\| = \mathbf{V}^{-1}$ are given in the Appendix.

Examples

The above procedure was applied to 60 items from an SAT mathematical section which had been given to 2691 examinees. The parameter estimates were obtained from a LOGIST run at ETS. Simultaneous 95% confidence limits for four of the 60 items are presented here.

The item response curve and corresponding confidence bands for Item 1 are given in Figure 1. The simultaneous confidence bands are similar to those found for simple linear regression lines in that they are quite narrow in the neighborhood of the mean of the distribution and flare out in the tails. The upper tails of the confidence bands are restricted by the asymptotic constraints particular to item response curves.

Insert Figures 1, 2, 3, and 4 about here

The effects of these constraints on the upper tail are clearly shown in Figure 2 which contains the results for Item 2. Here the confidence bands begin to widen around the upper portion of the response curve but are then forced back together due to the upper asymptotic constraint. In contrast, the lower portion of the confidence bands for this item appear to be reaching a constant width.

This behavior, which is particular to the lower tail of the response curve, is even more evident in the results for Item 3 which are presented in Figure 3. As the response curve depends primarily on the parameter c in the lower tail the variance contributing to the width of the confidence bands becomes increasingly dependent only on the variance related to $\hat{c}$.

The results for Item 4, contained in Figure 4, contrast those for Item 3 in that the lower tail band is still widening while the upper portion is fairly constant until the function constraints narrow the band. Of course, if the plotted range of abilities was increased enough all item confidence bands should show similar characteristics: Asymptotically constant about the lower tail with the width of the band determined by the variance associated with $\hat{c}$; narrowing about the mean of the distribution; first widening and then converging to the item response curve due to the asymptotic constraints in the upper tail.

Discussion

While standard errors for estimates of the three-parameter logistic item response model have been available to psychometricians, assuming abilities are known (see Lord, 1980, p. 191) precision information related to an entire response curve has not. Simultaneous confidence bands indicate the sampling variation of item response curves relative to a fitted function. This information is more relevant to psychometricians than standard errors corresponding to the parameter estimates as it provides for more insight into the behavior of item functions and hopefully their associated predictive limits.

The procedures outlined in this paper are simple enough computationally to be utilized economically in regular item response studies. Clearly, incorporating the calibration and graphical display of simultaneous confidence bands into calibration software would greatly benefit test designers.

References

Brand, R. J., Pinnock, D. E., & Jackson, K. L. (1973). Large sample confidence bands for the logistic response curve and its inverse. *American Statistician*, *27*, 157-160.

Hauck, W. W. (1983). A note on confidence bands for the logistic response curve. *American Statistician*, *37*, 158-160.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Thissen, D., & Wainer, H. (1983a). *Confidence envelopes for monotonic functions: Principals, derivations, and examples*. (AFHRL Technical Paper No. 82-37). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Thissen, D., & Wainer, H. (1983b). *Confidence envelopes for item theory*. (ETS Program Statistics Research Technical Report No. 83-38). Princeton, NJ: Educational Testing Service.

Wingersky, M. S. (1983). LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), *Applications of item response theory* (pp. 45-56). Vancover, BC: Educational Research Institute of British Columbia.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST user's guide*. Princeton, NJ: Educational Testing Service.

Appendix

The required values of $\|V^{rs}\| = V^{-1}$ are obtainable from the usual Fisher information matrix for a, b, and c (Lord, 1980, p. 191), as follows:

$$A = Da,\ B = -Dab ,$$

$$V^{11} = D^{-2}[I_{aa} + a^{-1}b(a^{-1}bI_{bb} - 2I_{ab})] ,$$

$$V^{12} = -D^{-2}a^{-1}(I_{ab} - a^{-1}bI_{bb}) ,$$

$$V^{22} = D^{-2}a^{-2}I_{bb} ,$$

$$V^{13} = D^{-1}(I_{ac} - a^{-1}bI_{bc}) ,$$

$$V^{23} = -D^{-1}a^{-1}I_{bc} ,$$

$$V^{33} = I_{cc} .$$

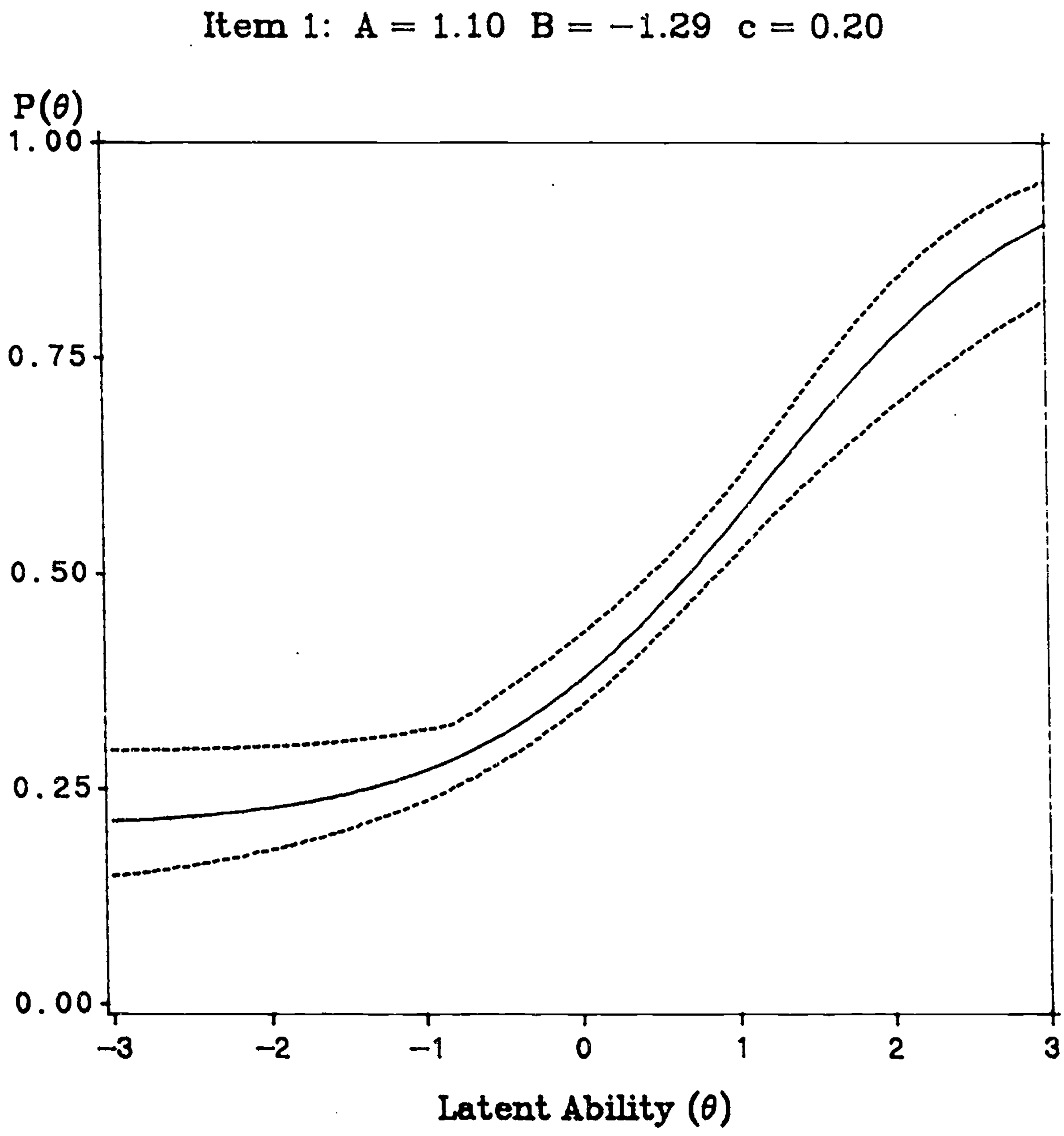

Figure 1.
Item response function and corresponding simultaneous confidence bands for Item 1 from an SAT mathematical section.

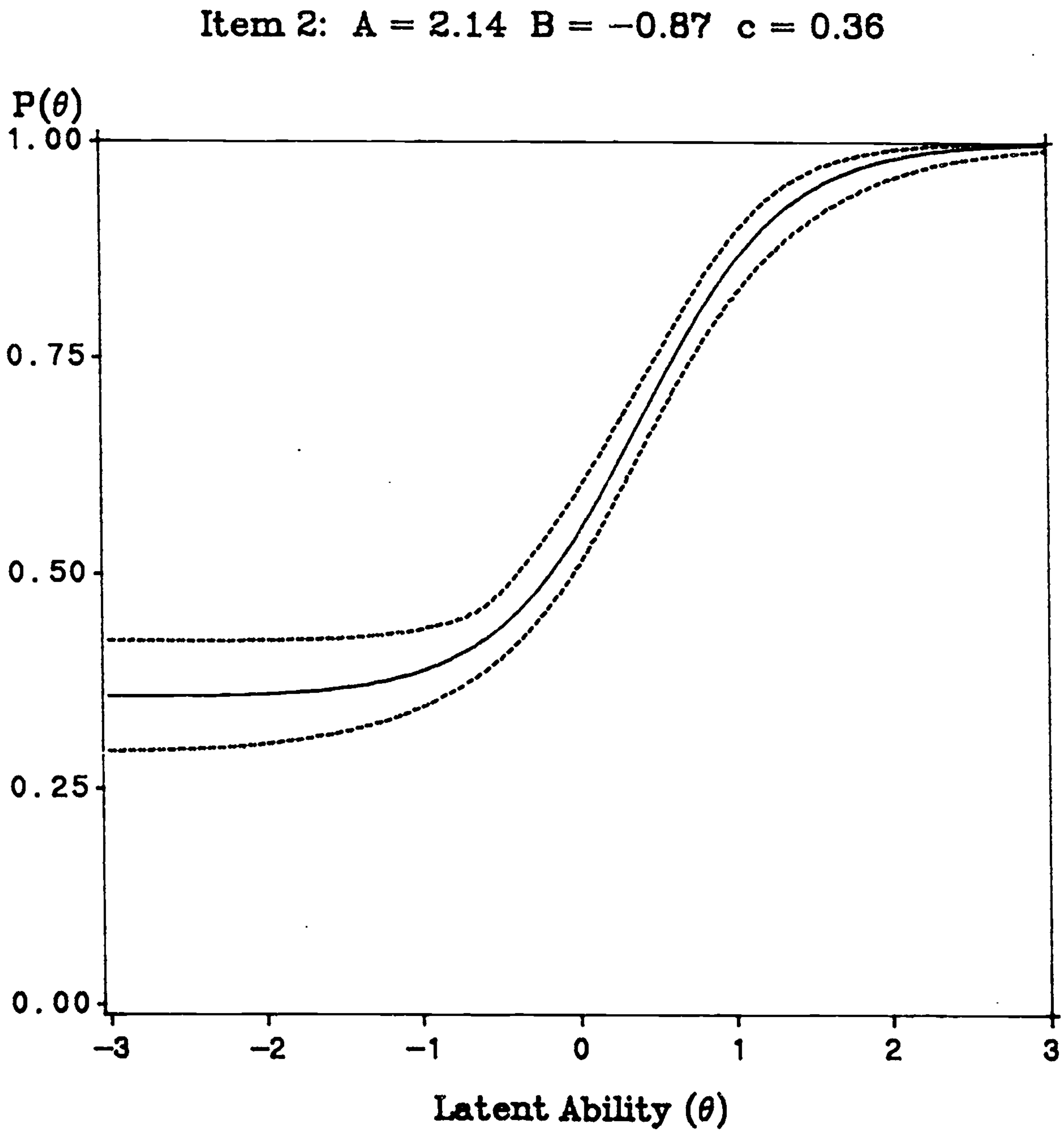

Figure 2.
Item response function and corresponding simultaneous confidence bands for Item 2 from an SAT mathematical section.

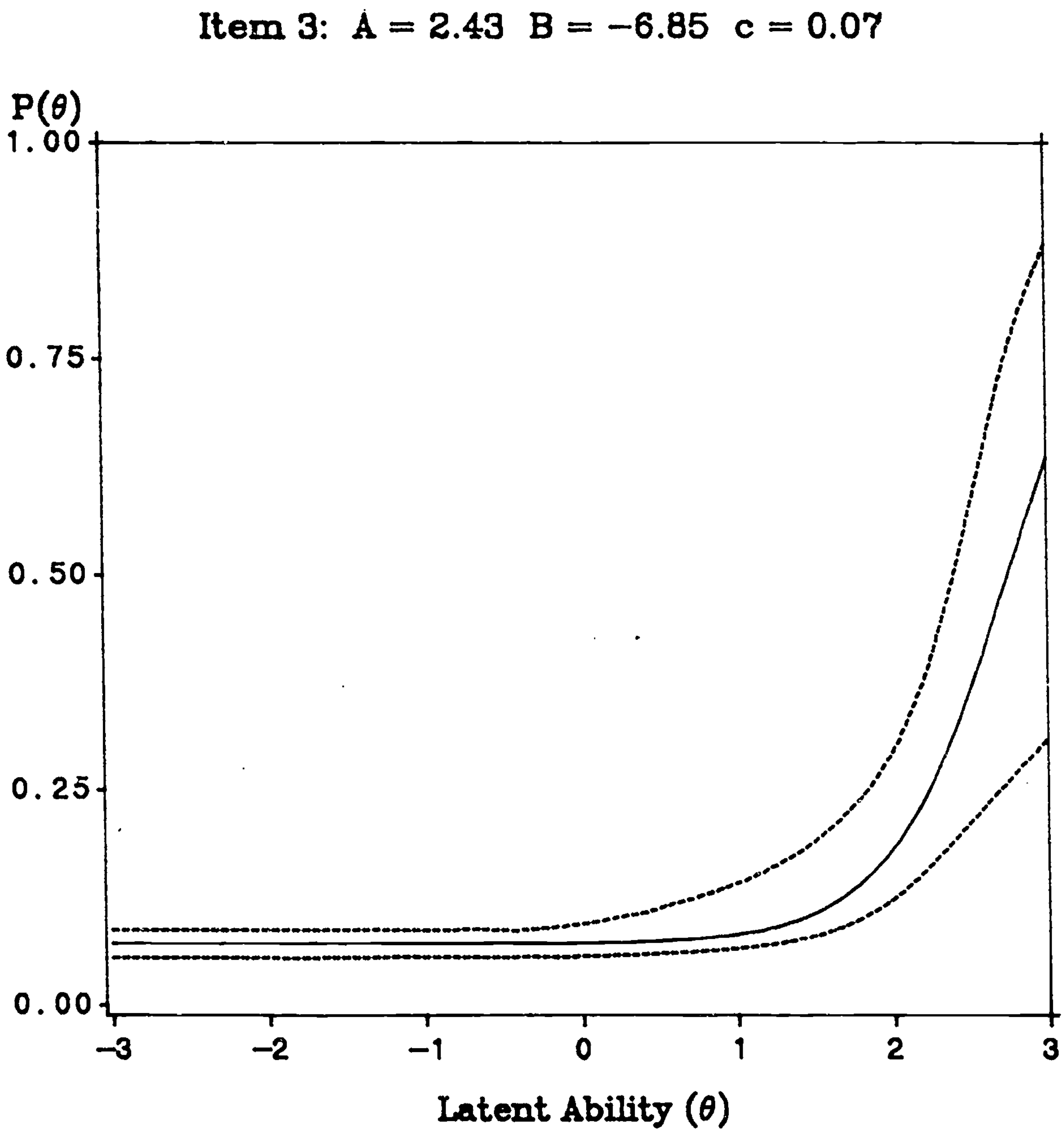

Figure 3.
Item response function and corresponding simultaneous confidence bands for Item 3 from an SAT mathematical section.

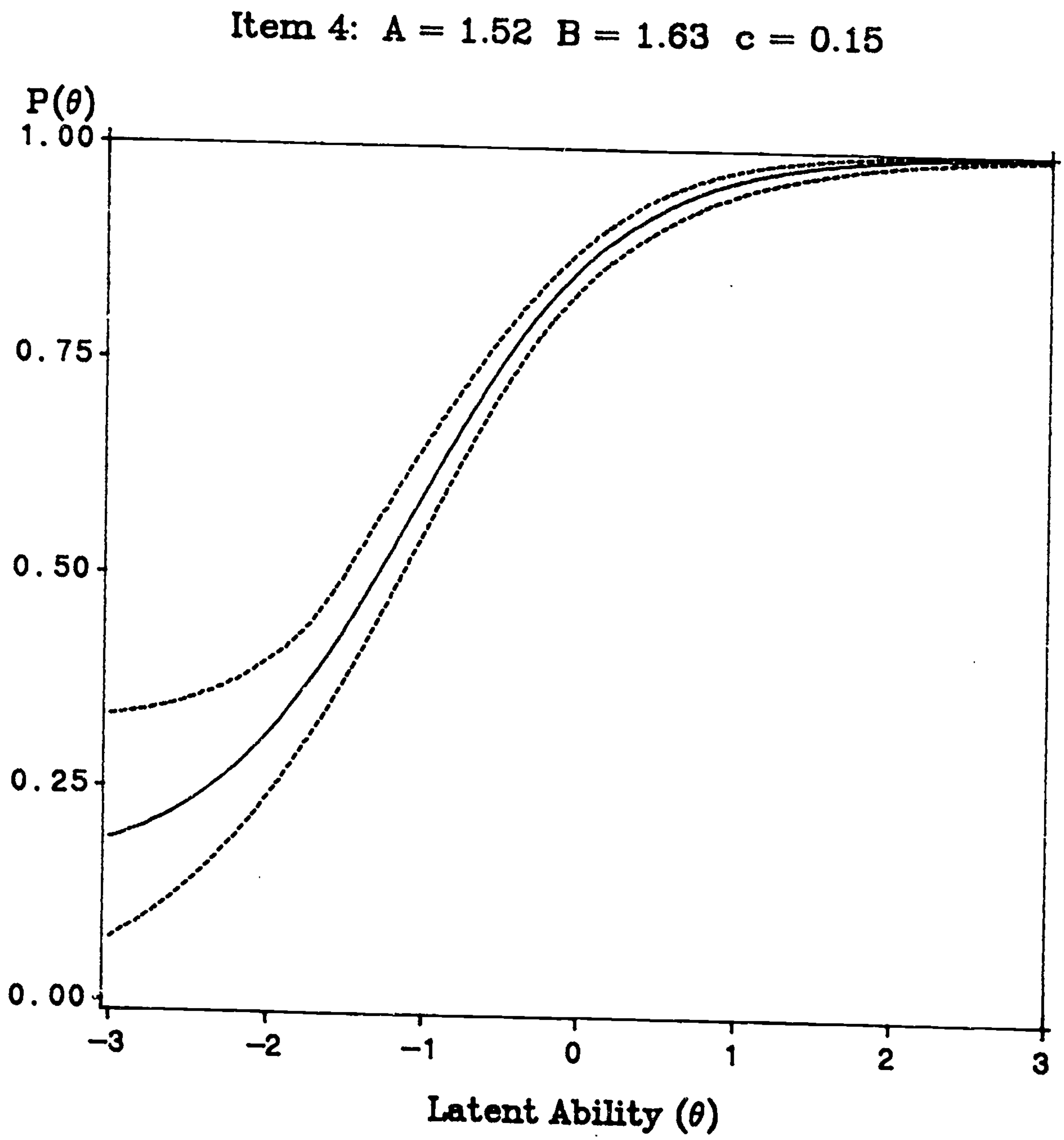

Figure 4.
Item response function and corresponding simultaneous confidence bands for Item 4 from an SAT mathematical section.